Discrete Applied Mathematics 4 (1982) 233-236
North-Holland Publishing Company

COMMUNICATION

A RESULT CONCERNING THE TRIOS GENERATED BY COMMUTATIVE SLIP-LANGUAGES

Juha KORTELAINEN

Department of Mathematics, University of Oulu, Oulu, Finland

Communicated by A Salomaa
Received 29 January 1982

1. Introduction

The trio generated by the commutative closure $c((a_1 \cdots a_k)^*)$ of the regular language $(a_1 \cdots a_k)^*$ is studied. For each positive integer k, let $\mathcal{R}_k$ be the set of all regular languages over k symbols, and let $c(\mathcal{R}_k)$ be the language family which consists of the commutative closures of all languages in $\mathcal{R}_k$. Latteux proves in [3] that the trio generated by the language family $c(\mathcal{R}_2)$ coincides with the trio generated by the language $c((a_1 a_2)^*)$, i.e., $\mathscr{C}(c(\mathcal{R}_2)) = \mathscr{C}(c((a_1 a_2)^*))$. In [4] it is asked whether or not the equality $\mathscr{C}(c(\mathcal{R}_k)) = \mathscr{C}(c((a_1 \cdots a_k)^*))$ holds for every $k \geq 3$. We shall give an affirmative answer to this question.

2. Preliminaries

The reader is referred to [1] for all concepts that are not defined here.

Let Σ_1 be a finite alphabet. For each x in Σ_1^* and each language $L \subseteq \Sigma_1^*$, define $c(x)$ to be the set of all words, which are permutations of the work x, and $c(L) = \bigcup \{c(x) \mid x \text{ in } L\}$. $c(L)$ is called the *commutative closure* of L. L is *commutative* if $L = c(L)$. For an a-transducer $M = (K_1, \Sigma_1, \Sigma_2, H, p_0, F)$, let f_i, $i = 1, 2, 3, 4$, be the homomorphisms on H^* defined as $f_1(t) = p$, $f_2(t) = x$, $f_3(t) = y$, and $f_4(t) = q$ for each $t = (p, x, y, q)$ in H. Let $M(L) = f_3(f_2^{-1}(L) \cap D_M)$, where D_M is the set of all computations of M. It should be noted that this definition and the definition of $M(L)$ used in [1] are equivalent.

Let $a_1, a_2, \ldots$ be different symbols. For each k in N_+, let $E_k = c((a_1 \cdots a_k)^*)$. Let Ψ be the Parikh-mapping from $\{a_1, \ldots, a_k\}^*$ into N^k. A language $L \subseteq \{a_1, \ldots, a_k\}^*$ is a SLIP-*language* if $\Psi(L)$ is a semilinear set.

For each language family $\mathscr{L}$, let $\mathscr{C}(\mathscr{L})$ ($\hat{\mathscr{C}}(\mathscr{L})$) be the smallest (full) trio containing $\mathscr{L}$.

0166-218X/82/0000-0000/$02.75 © 1982 North-Holland

3. The main theorem

We now prove our central result.

Theorem. *$\mathscr{C}(c(\mathscr{R}_k)) = \mathscr{C}(E_k)$ for each k in N_+.*

Proof. For $k=1$ we have $\mathscr{C}(c(\mathscr{R}_1)) = \mathscr{C}(E_1) = \mathscr{R}$.

Let $k \geq 2$ in N_+ be arbitrary but fixed.

Clearly $E_k = c((a_1 \cdots a_k)^*)$ is in $c(\mathscr{R}_k)$ and thus $\mathscr{C}(E_k) \subseteq \mathscr{C}(c(\mathscr{R}_k))$.

Consider the reverse inclusion. Let R be in $\mathscr{R}_k$. Since R is a SLIP-language (Lemma 1 of [2]) and $\mathscr{C}(E_k)$ is closed under e-free homomorphism and, as a principal trio, under union, we may without loss of generality assume that $c(R) = c(w_0 w_1^* \cdots w_p^*)$ for some p in N_+ and words $w_0, w_1, \ldots, w_p$ in $\{a_1, \ldots, a_k\}^*$. To prove that $c(R)$ is in $\mathscr{C}(E_k)$ it clearly suffices to show that the language $c(w_1^* \cdots w_p^*)$ is in $\mathscr{C}(E_k)$. For each j in $\{1, \ldots, p\}$, $\Psi(w_j) = (i_{j1}, \ldots, i_{jk})$ for some $i_{j1}, \ldots, i_{jk}$ in N. Let

$$L = \{a_1^{n_1 i_{11} + \cdots + n_p i_{p1}} \cdots a_k^{n_1 i_{1k} + \cdots + n_p i_{pk}} \mid n_1, \ldots, n_p \text{ in } N\}.$$

Since $c(L) = c(w_1^* \cdots w_p^*)$, it, by Corollary I.3 of [3], suffices to prove that L is in $\mathscr{C}(E_k)$. We define j_{rs} in N_+, $r = 1, \ldots, p$, $s = 1, \ldots, k$ as follows:

$$j_{rs} = \begin{cases} 1 & \text{if } i_{rs} = 0, \\ i_{rs} & \text{if } i_{rs} > 0. \end{cases}$$

For any set of elements $\{g_1, \ldots, g_m\} \subseteq N^k$, let $\mathscr{L}(g_1, \ldots, g_m)$ be the vector space over the field of real numbers generated by $g_1, \ldots, g_m$. For each h in N let d_h in N^k be defined as $d_h = (h, \ldots, h)$. Let $u_1, \ldots, u_p, v_1, \ldots, v_{k-1}$ in N^k be any elements such that

(i) $u_r, v_1, \ldots, v_{k-1}$ are linearly independent for each $r = 1, \ldots, p$;

(ii) d_1 is not in $\mathscr{L}(v_1, \ldots, v_{k-1})$; and

(iii) $j_{r1} u_r + j_{r2} v_1 + \cdots + j_{rk} v_{k-1} = d_{h_r}$ for some h_r in N_+, $r = 1, \ldots, p$.

It can be shown that such elements really exist. Let $x_1, \ldots, x_p, y_1, \ldots, y_{k-1}$ be the words in $a_1^* \cdots a_k^*$ for which $\Psi(x_r) = u_r$ and $\Psi(y_s) = v_s$, $r = 1, \ldots, p$, $s = 1, \ldots, k-1$. An a-transducer $M = (K_1, \Sigma_1, \Sigma_2, H, p_0, F)$ such that $L = M(E_k)$ is now constructed. Let

(1) $K = \{q_i, s_j \mid i = 0, \ldots, p, j = 1, \ldots, k-1\}$;

(2) $\Sigma_1 = \Sigma_2 = \{a_1, \ldots, a_k\}$;

(3) H consists of the following quadruples:

(a) (q_r, e, e, q_{r+1}), $r = 0, \ldots, p-1$, (q_p, e, e, s_1);

(b) $(q_r, x_r^{j_{r1}} z_{r1}^{j_{r2}} \cdots z_{r,k-1}^{j_{rk}}, a_1^{i_{r1}}, q_r)$ where $z_{rh} = y_h$ if $i_{r,h+1} = 0$ and $z_{rh} = e$ otherwise, $r = 1, \ldots, p$, $h = 1, \ldots, k-1$;

(c) (s_h, e, e, s_{h+1}), $h = 1, \ldots, k-2$, if $k > 2$; and

(4) $F = \{s_{k-1}\}$.

We briefly prove that $L = M(E_k)$.

Let w be in L. Then

$$w = a_1^{n_1 i_{11} + \cdots + n_p i_{p1}} \ldots a_k^{n_1 i_{1k} + \cdots + n_p i_{pk}}$$

for some $n_1, \ldots, n_p$ in N. Consider the computation

$$\begin{aligned} t = {} & (q_0, e, e, q_1)(q_1, x_1^{j_{11}} z_{11}^{j_{12}} \cdots z_{1,k-1}^{j_{1k}}, a_1^{i_{11}}, q_1)^{n_1} \cdots \\ & \cdots (q_{p-1}, e, e, q_p)(q_p, x_p^{j_{p1}} z_{p1}^{j_{p2}} \cdots z_{p,k-1}^{j_{pk}}, a_1^{i_{p1}}, q_p)^{n_p} (q_p, e, e, s_1) \\ & (s_1, y_1, a_2, s_1)^{n_1 i_{12} + \cdots + n_p i_{p2}} \ldots (s_{k-2}, e, e, s_{k-1}) \\ & (s_{k-1}, y_{k-1}, a_k, s_{k-1})^{n_1 i_{1k} + \cdots + n_p i_{pk}}. \end{aligned}$$

It is not difficult to see that $w = f_3(t)$ and that $f_2(t)$ is in E_k. Thus w is in $M(E_k)$.

Assume now that w is in $M(E_k)$. Then there must be a computation

$$\begin{aligned} t' = {} & (q_0, e, e, q_1)(q_1, x_1^{j_{11}} z_{11}^{j_{12}} \cdots z_{1,k-1}^{j_{1k}}, a_1^{i_{11}}, q_1) \cdots \\ & \cdots (q_{p-1}, e, e, q_p)(q_p, x_p^{j_{p1}} z_{p1}^{j_{p2}} \cdots z_{p,k-1}^{j_{pk}}, a_1^{i_{p1}}, q_p)^{n_p} (q_p, e, e, s_1) \\ & (s_1, y_1, a_2, s_1)^{m_1} \cdots (s_{k-2}, e, e, s_{k-1})(s_{k-1}, y_{k-1}, a_k, s_{k-1})^{m_{k-1}}, \end{aligned}$$

$n_1, \ldots, n_k, m_1, \ldots, m_{k-1}$ in N such that $w = f_3(t')$ and $f_2(t')$ is in E_k. On the other hand

$$\begin{aligned} t = {} & (q_0, e, e, q_1)(q_1, x_1^{j_{11}} z_{11}^{j_{12}} \cdots z_{1,k-1}^{j_{1k}}, a_1^{i_{11}}, q_1)^{n_1} \cdots \\ & \cdots (q_{p-1}, e, e, q_p)(q_p, x_p^{j_{p1}} z_{p1}^{j_{p2}} \cdots z_{p,k-1}^{j_{pk}}, a_1^{i_{pk}}, q_p)^{n_p} \\ & (q_p, e, e, s_1)(s_1, y_1, a_2, s_1)^{n_1 i_{12} + \cdots + n_p i_{p2}} \ldots (s_{k-2}, e, e, s_{k-1}) \\ & (s_{k-1}, y_{k-1}, a_k, s_{k-1})^{n_1 i_{1k} + \cdots + n_p i_{pk}} \end{aligned}$$

is a computation such that $f_2(t)$ is in E_k. Basing on the Parikh-mappings of the words $f_2(t')$ and $f_2(t)$ and on the conditions (i), (ii) and (iii) it can be shown that $t = t'$, and thus

$$w = f_3(t') = f_3(t) = a_1^{n_1 i_{11} + \cdots + n_p i_{p1}} \ldots a_k^{n_1 i_{1k} + \cdots + n_p i_{pk}}$$

is in L.

Now, by Theorem 3.2.1. of [1], $L = M(E_k)$ is in $\hat{\mathscr{C}}(E_k)$. Clearly E_k is a SLIP-language and thus, by Proposition II.11 of [3], the family $\mathscr{C}(E_k)$ is closed under arbitrary homomorphism, i.e., $\mathscr{C}(E_k) = \hat{\mathscr{C}}(E_k)$. So $L = M(E_k)$ is in $\mathscr{C}(E_k)$ and the proof is complete. □

The previous theorem allows us to answer a question of Latteux in [4]:

Corollary. *For each positive integer k, the language family* $\mathscr{C}(c(R_k))$ *is a (full) principal trio.*

It should be noted that by the application of AFL-techniques used in [1], it can be shown that the smallest AFL containing the language family $c(R_k)$ is a (full) principal AFL.

Note. Let $b_1, b_2, \ldots$ be new symbols. Define $O_k = c((a_1 b_1)^* \cdots (a_k b_k)^*)$. Using a proof which resembles the proof of the theorem above, it can be shown that O_k is in $\mathscr{C}(E_{k+1})$. Then it is easily seen that $\mathscr{C}(E_{k+1}) = \mathscr{C}(c(\mathscr{R}_{k+1})) = \mathscr{C}(O_k) = \mathscr{H}(\mathscr{L}_1 \wedge \cdots \wedge \mathscr{L}_k)$ where $\mathscr{L}_i = \mathscr{C}(E_2)$, $i = 1, \ldots, k$, k in N_+. All the trios above are full.

Acknowledgment

I am very grateful to Professor Paavo Turakainen for encouragement and introducing me to the problem first raised by M. Latteux.

References

[1] S. Ginsburg, Algebraic and Automata-Theoretic Properties of Formal Languages (North-Holland, Amsterdam, 1975).
[2] M. Latteux, Cônes rationnels commutativement clos, RAIRO Inform. Théor. 11 (1977) 29–51.
[3] M. Latteux, Cônes rationnels commutatifs, J. Comput. System Sci. 18 (1979) 307–333.
[4] M. Latteux, Langages commutatifs, transductions rationnelles et intersection, Publication l'Equipe Lilloise d'Informatique Théorique, IT 34.81 (1981).